Appl. Math. Lett. Vol. 1, No. 3, pp. 211-216, 1988
Printed in Great Britain
0893-9659/88 $3.00 + 0.00
Pergamon Press plc

Parameter Estimation in Nonlinear Distributed Systems-Approximation Theory and Convergence Results

H.T. BANKS[1], SIMEON REICH[2] and I.G. ROSEN[3]

[1]Center for Control Sciences, Division of Applied Mathematics, Brown University, Providence, RI 02912, U.S.A.
[2]Department of Mathematics, University of Southern California, Los Angeles, CA 90089, U.S.A. and
Department of Mathematics, The Technion-Israel Institute of Technology, Haifa 32000, Israel
and [3]Department of Mathematics, University of Southern California, Los Angeles, CA 90089, U.S.A.

1. INTRODUCTION

In this short paper we describe an abstract approximation framework and convergence theory for Galerkin approximations applied to inverse problems involving nonlinear distributed parameter systems. We consider parameter estimation problems formulated as the minimization of a least-squares like performance index over a compact admissible parameter set subject to state constraints given by an inhomogeneous nonlinear distributed system. Our theory applies to systems whose dynamics can be described by either time independent or nonstationary strongly maximal monotone operators defined on a reflexive Banach space which is densely and continuously embedded in a Hilbert space. Our approach involves the use of standard Galerkin techniques to obtain a sequence of approximating identification problems, each of which involves finite dimensional state constraints. We demonstrate that if readily verifiable conditions on the system's dependence on the unknown parameters are satisfied, and the usual Galerkin approximation assumption (i.e. strong convergence of the orthogonal projections in a sufficiently strong norm topology) holds, then solutions to the approximating problems exist and in some sense approximate a solution to the original infinite dimensional identification problem. These results also guarantee a type of stability - i.e. continuous dependence of approximate estimates on the observations.

In section 2 we summarize the theory developed in [2] for the case of a temporally homogeneous operator, while in Section 3 we treat systems involving nonstationary operators. The details for the latter case can be found in [3]. Our results here completely subsume the linear theory developed in [1] which is applicable to inverse problems for abstract parabolic systems involving regularly dissipative or strongly elliptic linear operators induced by bounded, coercive, sesquilinear forms. In the fourth section we briefly mention some examples, applications, and preliminary numerical results.

2. THE TEMPORALLY HOMOGENEOUS CASE

Let $\{H, <\cdot,\cdot>, |\cdot|\}$ be a real Hilbert space and let $\{V, \|\cdot\|\}$ be a reflexive Banach space. We assume that V is densely and continuously embedded in H and consequently it follows that $V \hookrightarrow H \hookrightarrow V^*$ with H densely and continuously embedded in V^*. Let $\mathcal{Q}$ and Z be metric spaces and let Q be a nonempty, sequentially compact subset of $\mathcal{Q}$. We consider inverse problems of the following general form

(ID) Given observations $z \varepsilon Z$, determine parameters $q \varepsilon Q$ which minimize the functional

$$\phi(q) = \Phi(u(q); z)$$

Typeset by $\mathcal{A}_{\mathcal{M}}\mathcal{S}$-TEX where u(q) = u(·;q) is the mild solution to the initial value problem

$$\dot{u}(t) + A(q)u(t) = f(t;q), \qquad 0 \le t \le T \tag{2.1}$$

$$u(0) = u_0(q). \tag{2.2}$$

In problem (ID) above we assume that $T > 0$, that for each $z \varepsilon Z$, $\Phi(\cdot;z)$ is a continuous map from C(0,T;H) into $BbbR^+$, and that for each $q\varepsilon Q$, A(q) is a single valued, hemicontinuous, everywhere defined, in general nonlinear, operator from V into $V^*, f(\cdot;q)\varepsilon L_1(0,T;H)$, and $u_0(q)\varepsilon H$. In addition, we require the following assumptions.
(A) (Continuity) For each $v\varepsilon V$ the map $q \to A(q)v$ is continuous from $Q \subset \mathcal{Q}$ into V^* and the
mappings $q \to f(t;q)$ for almost every $t\varepsilon[0,T]$, and $q \to u_0(q)$, are continuous from $Q \subset \mathcal{Q}$
into H.
(B) (Equi-V-monotonicity) There exist an $\omega\varepsilon BbbR$ and an $\alpha > 0$, both independent of $q\varepsilon Q$ such that

$$< A(q)u - A(q)v, u - v > +\omega|u-v|^2 \ge \alpha \parallel u - v \parallel^2, \quad for \quad every \quad u, v\varepsilon V.$$

(C) (Equi-boundedness) There exists a constant $\beta > 0$, independent of $q\varepsilon Q$ for which

$$\parallel A(q)v \parallel_* \le \beta(\parallel v \parallel + 1) \quad for \quad every \quad v\varepsilon V.$$

In (B) above $< \cdot,\cdot >$ is understood to be the usual extension of the H inner product to the duality pairing between V and V^*, and in (C), $\parallel \cdot \parallel_*$ denotes the usual operator norm on V^*. If for each $q\varepsilon Q$ we define the operator $A_0(q) : Dom(A_0(q)) \subset H \to H$ to be the restriction of the operator A(q) to the set $Dom(A_0(q)) = \{v\varepsilon V : A(q)v\varepsilon H\}$, then the assumptions (A) - (C) imply (see [2]) that there exists a unique nonlinear evolution system, $\{U(t,s;q) : 0 \le s \le t \le T\}$, on H, corresponding to the operator $A_0(q)$. The mild solution to the initial value problem (2.1), (2.2) is then defined to be $u(t;q) = U(t,0;q)u_0(q), \quad 0 \le t \le T$.
We develop computational methods for the solution of problem (ID) by constructing a sequence of finite dimensional approximating identification problems. We do this via a Galerkin approach. For each n = 1,2,... let H_n denote a finite dimensional subspace of H with $H_n \subset V$, and let $P_n : H \to H_n$ denote the orthogonal projection of H onto H_n. We require the assumption that
(D) $\lim_{n\to\infty} \parallel P_n v - v \parallel = 0$ for each $v\varepsilon V$.
Note that assumption (D) also implies that $\lim_{n\to\infty} |P_n u - u| = 0$, for each $u\varepsilon H$. We define Galerkin approximations, $A_n(q) : H_n \to H_n$, to A(q) by setting $A_n(q)u_n = v_n$ where v_n is that unique element in H_n (guaranteed to exist by the Riesz Representation Theorem) which satisfies $< A(q)u_n, w_n > = < v_n, w_n >, w_n\varepsilon H_n$. We define $f_n(\cdot;q)\varepsilon L_1(0,T;H_n)$ and $u_{0n}(q)\varepsilon H_n$ by $f_n(t;q) = P_n f(t;q)$, for almost every $t\varepsilon[0,T]$ and $u_{0n}(q) = P_n u_0(q)$. We consider the following sequence of identification problems.
(ID_n) Given observations $z\varepsilon Z$, determine parameters $\overline{q}_n\varepsilon Q$ which minimize the functional

$$\phi_n(q) = \Phi(u_n(q);z)$$

where $u_n(q) = u_n(\cdot;q)$ is the mild solution to the initial value problem

$$\dot{u}_n(t) + A_n(q)u_n(t) = f_n(t;q), \quad 0 < t \le T \tag{2.3}$$

$$u_n(0) = u_{0n}(q). \tag{2.4}$$

The definition of the operators $A_n(q)$, and assumptions (B) and (C) imply that for each $q\varepsilon Q$ $A_n(q)$ is m-accretive on H_n (see [4]). It follows that there exists a unique nonlinear evolution system $\{U_n(t,s;q) : 0 \le s \le t \le T\}$ on H_n corresponding to $A_n(q)$. The mild solution to the initial value problem (2.3), (2.4) is then given by $u_n(t;q) = U_n(t,0;q)u_{0n}(q)$ for $0 \le t \le T$.

Assume for the moment that (i) each of the identification problems (ID_n) has a solution $\overline{q}_n \varepsilon Q$. Then under the further condition that (ii) for any sequence $\{q_n\} \subset Q$ with $\lim_{n\to\infty} q_n = q\varepsilon Q$ we have $\lim_{n\to\infty} u_n(q_n) = u(q)$ in $C(0,T;H)$, it follows that solutions to $(ID_n), \overline{q}_n$, approximate a solution $\overline{q}$ to problem (ID) in the sense that $\overline{q}$ exists as a subsequential limit of the sequence $\{\overline{q}_n\}$. Indeed, $\{\overline{q}_n\} \subset Q$, and Q compact imply there exists a subsequence $\{\overline{q}_{n_j}\} \subset \{\overline{q}_n\}$ with $\lim_{n\to\infty} \overline{q}_{n_j} = \overline{q}\varepsilon Q$. Then for any $q\varepsilon Q$, condition (ii) and the continuity of Φ imply

$$\phi(\overline{q}) = \Phi(\lim_{j\to\infty} u_{n_j}(\overline{q}_{n_j}); z) = \lim_{j\to\infty} \phi_{n_j}(\overline{q}_{n_j}) \le \lim_{j\to\infty} \phi_{n_j}(q) = \Phi(\lim_{j\to\infty} u_{n_j}(q); z) = \phi(q).$$

We note that it was not necessary to assume that a solution to problem (ID) exists, but only that solutions to the approximating identification problems (ID_n) exist. Also, when there is a unique solution to problem (ID), it is clear that the sequence $\{\overline{q}_n\}$ itself will converge to $\overline{q}$.

By applying a nonlinear analog of the familiar Trotter-Kato approximation result for linear semigroups of operators we are able to show that our assumptions (A) - (D) are in fact sufficient to conclude that conditions (i) and (ii) above are satisfied. Indeed, condition (i) will follow from the compactness of Q, the continuity of Φ, and the continuous dependence result

$$\lim_{m\to\infty} u_n(q_m) = u_n(q) \quad in \quad C(0,T;H_n) \tag{2.5}$$

for each n, whenever $\{q_m\} \subset Q$ with $\lim_{m\to\infty} q_m = q$. As in the linear case, the nonlinear approximation result is based upon convergence of the resolvent (see [2]). We are able to demonstrate that assumptions (A) - (D) imply $\lim_{n\to\infty}(I + \lambda A_n(q_n))^{-1} P_n \varphi = (I + \lambda A(q))^{-1}\varphi$, and $\lim_{m\to\infty}(I+\lambda A_n(q_m))^{-1}\varphi_n = (I+\lambda A_n(q))^{-1}\varphi_n$ in H for each $\varphi\varepsilon H$ and $\varphi_n \varepsilon H_n$ and some $\lambda > 0$ with $\lambda\omega < 1$ whenever $\lim_{n\to\infty} q_n = q$ and $\lim_{m\to\infty} q_m = q$. This together with the continuity assumptions on f and u_0 given in (A), and the definitions of f_n and u_{0n} yield the desired convergence in (2.5) and condition (ii).

3. THE TEMPORALLY INHOMOGENEOUS CASE

We consider briefly the case in which the operator A(q) depends on time. That is, A(q) = A(t;q), and the parameters q are permitted to be time dependent as well. If we add the assumption

(E)(Measurability) For each $q\varepsilon Q$ the function $t \to A(t;q)v(t) : [0,T] \to V^*$ is strongly measurable

for every $v\varepsilon L_2(0,T;V)$,

and require that assumptions (A) - (C) with A(q) replaced by A(t;q) are satisfied for almost every $t\varepsilon[0,T]$, then the results outlined in the previous section continue to hold. We note however, that the general approach and techniques used to obtained these results are in fact quite different. While there do exist analogs of the generic Trotter-Kato like approximation result referred to earlier which are applicable in the time dependent case, their application requires that a far stronger and far more difficult to verify (for both the original system and the Galerkin approximations) condition be satisfied than assumption

(E) above (see [2]). Instead we argue convergence directly via variational arguments in the context of an operator theoretic formulation suggested by Barbu [4].

We briefly outline the essential features of our approach here. The details can be found in [3]. Define the reflexive Banach space $\mathcal{V}$ and the Hilbert space $\mathcal{H}$ by $\mathcal{V} = L_2(0,T;V)$ and $\mathcal{H} = L_2(0,T;H)$, respectively. Then $\mathcal{V}^* = L_2(0,T;V^*)$ and $\mathcal{V} \hookrightarrow \mathcal{H} \hookrightarrow \mathcal{V}^*$. For each $q\varepsilon Q$ define the operator $\mathcal{B}(q) : Dom(\mathcal{B}(q)) \subset \mathcal{V} \to \mathcal{V}^*$ by $\mathcal{B}(q)v = \dot{v}$ for $v\varepsilon Dom(\mathcal{B}(q)) = \{w\varepsilon\mathcal{V} : \dot{w}\varepsilon\mathcal{V}^*, w(0) = u_0(q)\}$. The derivatives in the above definition are assumed to be in a generalized or distributional sense. Define the operator $\mathcal{A}(q) : \mathcal{V} \to \mathcal{V}^*$ by $(\mathcal{A}(q)v)(t) = A(t;q)v(t)$, a.e. $t\varepsilon[0,T]$. It is not difficult to show that (see [4]) $\mathcal{B}(q)$ is maximal monotone in $\mathcal{V} \times \mathcal{V}^*$ and that $\mathcal{A}(q)$ is monotone, everywhere defined and hemicontinuous. If we define $\mathcal{T}(q) = \mathcal{A}(q) + \mathcal{B}(q) : Dom(\mathcal{B}(q)) \subset \mathcal{V} \to \mathcal{V}^*$, it follows that $\mathcal{R}(\mathcal{T}(q)) = \mathcal{V}^*$, and that the unique solution to the initial value problem (2.1), (2.2) (with A(q) replaced by A(t;q)), $u(q)\varepsilon L_2(0,T;V) \cap C(0,T;H) \cap H^1(0,T;V^*)$ is given by $u(q) = \mathcal{T}(q)^{-1}f(q)$ where $f(q) = f(\cdot;q)\varepsilon\mathcal{V}^*$.

With regard to approximation, we let H_n and P_n be defined as they were in Section 2, and let the definition of $A_n(t;q)$ be analogous to the definition of $A_n(q)$ given above. For each n = 1,2,... let $\mathcal{H}_n = L_2(0,T;H_n)$ and define $\mathcal{B}_n(q) : Dom(\mathcal{B}_n(q)) \subset \mathcal{H}_n \to \mathcal{H}_n$ by $\mathcal{B}_n(q)v_n = \dot{v}_n$ for $v_n\varepsilon Dom(\mathcal{B}_n(q)) = \{w_n\varepsilon\mathcal{H}_n : \dot{w}_n\varepsilon\mathcal{H}_n, w_n(0) = P_n u_0(q)\}$. Define $\mathcal{A}_n(q) : \mathcal{H}_n \to \mathcal{H}_n$ by $(\mathcal{A}_n(q)v_n)(t) = A_n(t;q)v_n(t)$ for $v_n\varepsilon\mathcal{H}_n$ and a.e. $t\varepsilon[0,T]$. If we set $\mathcal{T}_n(q) = \mathcal{A}_n(q) + \mathcal{B}_n(q)$, then once again we have $\mathcal{R}(\mathcal{T}_n(q)) = \mathcal{H}_n$ and the solution $u_n(q)$ to the initial value problem (2.3), (2.4) is given by $u_n(q) = \mathcal{T}_n(q)^{-1}P_n f(q)$. It can be shown directly that $\mathcal{T}_n(q_n)^{-1}P_n f(q_n) \to \mathcal{T}(q)^{-1}f(q)$ in C(0,T;H) as $n \to \infty$ whenever $\lim q_n = q$ and that $\mathcal{T}_n(q_m)^{-1}P_n f(q_m) \to \mathcal{T}_n(q)^{-1}P_n f(q)$ in C(0,T;H) as $m \to \infty$ whenever $\lim q_m = q$. Thus (i) and (ii) are satisfied and the desired result concerning the approximation of a solution $\overline{q}$ to problem (ID) by solutions $\overline{q}_n$ to the problems (ID_n) holds.

When the space V is separable (as is usually the case in practice) weak and strong measurability are equivalent. Consequently when this is the case, assumption (E) can be verified by showing that the complex valued function $t \to \langle A(t;q)u, v \rangle$ is measurable on [0,T] for all u, $v\varepsilon V$.

4. EXAMPLES, APPLICATIONS, AND NUMERICAL RESULTS

The theory presented here compeletely subsumes the linear theory given in [1]. In particular therefore, our results apply to abstract parabolic systems - i.e. those whose dynamics can be described by regularly dissipative linear operators, A(q) or A(t;q), induced by bounded, coercive, sesquilinear forms on $V \times V$. This class includes strongly elliptic partial differential operators on Sobolev spaces.

More generally, our theory can be applied to a wide class of systems involving so called nonlinear elliptic operators. Let Ω be a bounded region in R^ℓ with smooth boundary. Let m be a fixed nonnegative integer and set $N = \sum_{j=0}^{m} \ell^j$. Let $\mathcal{Q} = \times L_\infty((0,T) \times \Omega \times R^N)$ where the cartesian product is taken over all ℓ dimensional multi-indices α and β with $0 \leq |\alpha|, |\beta| \leq m$. Let Q be a compact subset of $\mathcal{Q}$ with the property that $q = \{q_{\alpha,\beta}\}, 0 \leq |\alpha|, |\beta| \leq m$ is an element in Q if (1) the mappings $\xi \to q_{\alpha,\beta}(t,x,\xi)$ are continuous for almost all $(t,x)\varepsilon[0,T] \times \Omega$ and α, β with $0 \leq |\alpha|, |\beta| \leq m$, and (2) there exists a positive constant λ, independent of $q\varepsilon Q$ for which

$$\sum_{0 \leq |\alpha|,|\beta| \leq m} (q_{\alpha,\beta}(t,x,\xi)\xi_\beta - q_{\alpha,\beta}(t,x,\eta)\eta_\beta)(\xi_\alpha - \eta_\alpha) \geq \lambda \sum_{0 \leq |\alpha| \leq m} |\xi_\alpha - \eta_\alpha|^2$$

for almost all $(t,x)\varepsilon[0,T] \times \Omega$ and all $\xi, \eta\varepsilon R^N$. We take $H = L_2(\Omega)$ and assume that V is a closed subspace of $H^m(\Omega)$ which contains $H_0^m(\Omega)$. For each $q = q_{\alpha,\beta}\varepsilon Q$ and almost every $t\varepsilon[0,T]$ we consider the nonlinear elliptic operator $A(t;q): V \to V^*$ given by

$$< A(t;q)u, v >= \sum_{0\le|\alpha|,|\beta|\le m} \int_\Omega q_{\alpha,\beta}(t,x,\delta u(x))D^\beta u(x) D^\alpha v(x) dx$$

where δu denotes the N-vector valued function whose components are $D^\alpha u$ for all multi-indices α with $0 \le |\alpha| \le m$. With the operator A(t;q) as defined above, the equation (2.1) corresponds to the quasilinear parabolic partial differential equation

$$\frac{\partial u}{\partial t}(t,x) + \sum_{0\le|\alpha|,|\beta|\le m} (-1)^{|\alpha|} D^\alpha q_{\alpha,\beta}(t,x,\delta u(t,x)) D^\beta u(t,x) = f(t,x;q), a.e. x\varepsilon\Omega, t > 0$$

together with boundary conditions appropriate to the choice of the space V (for example, Dirichlet, if $V = H_0^m(\Omega)$,Neumann, if $V = H^m(\Omega)$, etc.).

For the class of systems that we have just described, it is not difficult to show that assumptions (A) - (C), and (E) are satisfied. A number of applications, including for example, size structured and density dependent population dynamics, biological mixing in deep sea sediment cores, and nonlinear heat conduction/mass transfer, lead to inverse problems involving systems of this general type, and consequently therefore, ones to which our theory would apply (see [2] and [3] for details). With regard to approximation, most standard finite element schemes, for example, those based upon polynomial or Hermite spline, modal, or spectral (involving, for example, orthogonal polynomials) approximation, can be shown to satisfy assumption (D).

In practice, it is frequently the case that the parameter space $\mathcal{Q}$ and the admissible parameter set Q are infinite dimensional with elements consisting of spatially and/or temporally varying functions. When this is the case, solving the approximating identification problems (ID_n) requires that the admissible parameter set Q also be discretized. Briefly, this can be carried out as follows. For each m = 1,2,... Let $I^m : Q \subset \mathcal{Q} \to \mathcal{Q}$ be a continuous map with finite dimensional range and the property that $\lim_{m\to\infty} I^m(q) = q$, uniformly in q for $q\varepsilon Q$. We set $Q^m = I^m(Q)$ (note that Q^m is compact) and consider the doubly indexed sequence of approximating identification problems (ID_n^m) given by problem (ID_n) with Q replaced by Q^m. It can be shown that each of the problems admits a solution $\overline{q}_n^m \varepsilon Q^m$ and that there exists a subsequence $\{\overline{q}_{n_k}^{m_j}\} \subset \{\overline{q}_n^m\}$ for which $\{\overline{q}_{n_k}^{m_j}\} \to \overline{q}$ as $j,k \to \infty$, with $\overline{q}$ a solution to problem (ID). Once bases for H_n and the range of I^m have been chosen, the problem (ID_n^m) becomes one involving the minimization of a continuous functional over a closed and bounded subset of Euclidean space subject to finite dimensional (i.e. ODE) constraints. The resulting optimization problems are typically solved using an iterative search procedure (i.e. gradient, conjugate gradient, steepest descent, Newton's methods, etc.), a variety of which have been implemented as a part of any one of a number of standard, readily available software packages (for example, IMSL, MINPACK,etc.).

Finally we note that we have carried out some preliminary numerical studies on a Cray supercomputer wherein we used our approach together with a spline based approximation sheme to estimate a functional parameter in a quasilinear model for one dimensional heat conduction. With this method we were able to successfully identify the thermal conductivity which was assumed to be a function of the temperature gradient. To date, all of our studies have involved simulation data only. We intend to test our schemes on more sophisticated examples, some of which will involve the use of actual experimental data, and then to report our findings in detail in forthcoming papers.

Acknowledgements The first author was supported in part under grants NSF-MCS-8504316,
NASA-NAG-1-517, AFOSR-84-0398 and AFOSR-F49620-86-C-0111. The second author was supported in part by The Fund for the Promotion of Research at the Technion and by The Technion VPR Fund. The third author was supported in part under grants AFOSR-84-0393 and AFOSR-87-0356. Part of this research was carried out while the first and third authors were visiting scientists at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA, which is operated under NASA contracts NAS1-17070 and NAS1-18107.

REFERENCES

1. H.T. Banks and K. Ito, *A unified framework for approximation and inverse problems for distributed parameter systems*, Control Theory and Advanced Technology 4 (1988), 73–90.
2. H.T. Banks, S. Reich , and I.G. Rosen, *An approximation theory for the identification of nonlinear distributed parameter systems*, SIAM J. Contr. and Opt., submitted.
3. H.T. Banks, S. Reich, and I.G. Rosen, *Galerkin approximation for inverse problems for nonautonomous nonlinear systems*, Appl. Math. Opt., submitted.
4. V. Barbu, "Nonlinear Semigroups and Differential Equations in Banach Spaces," Noordhoff International, Leyden, The Netherlands, 1976.